I. Summary[1]

The ZIP file, located at `https://reopeningdata.github.io/` , contains the replication materials for "Reopening the Economy: What Are the Risks, and What Have States Done?" by Enghin Atalay, Shigeru Fujita, Sreyas Mahadevan, Ryan Michaels, and Tal Roded.

The two main do files that need to be run, in order, are `create_dataset.do` and `reopening_figures.do`. The first of the two files constructs the main datasets (`reopening_dataset.dta` and `reopening_dataset_combined.dta`). The second of the two do files produces the figures and tables that are included in the note and accompanying website.

II. On Measuring States' Initial Designation of Essential Industries

The appendix of "Reopening the Economy: What Are the Risks, and What Have States Done?" provides some background on our procedure to construct an industry-by-county dataset of closures (as of April 17, 2020). Here, we review this material, referring specifically to the replication files.

We follow two different procedures, one for a set of states (and counties) with relatively concise closing orders, and a second for states with lengthier orders. For the states with more concise orders, we hand coded the NAICS industries that we identify as closed or open. Our hand-coded industries are collected in `essential_industries_initial.xlsx`. For states with lengthier orders, we develop a list of keywords (collected in `finalkey.csv`) associated with each NAICS industry. Then, we search for these keywords in the text of states' and counties' closing orders. The output of this searching procedure is collected in `initial_other_states_based_on_cisa3.csv`. The code (and raw text from states' orders) to produce `itial_other_states_based_on_cisa3.csv` is given in the folder Create_Keyword_Based_Designations. Using `essential_industries_initial.xlsx` and `initial_other_states_based_on_cisa3.csv` as inputs, the code to produce a STATA database of states' initial closures is given in step 1 of `create_dataset.do`.

III. On Measuring Reopening

We follow a two-step procedure to construct our dataset of reopening paths by industry and county. The first step involves hand collecting information from states' reopening orders, storing this information in excel files. The second step, executed in STATA, involves compiling this information into a structured dataset; in this dataset, observations are NAICS industry-county pairs and variables indicate whether the county-industry pair should be marked as "open" or "closed" on a given week (for each week between May 1 and the most recently available period.) We describe the two steps in turn.

III.A Collecting Information from States' Reopening Orders

[1] Research results and conclusions expressed throughout this project are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia, the Federal Reserve System, or the Federal Reserve Board of Governors.

Our data collection efforts center on reading through states' executive orders and local news stories.

There are two groups of states in this step. First, certain states have clearly defined "phases" and within-state regions, whereby the movement into successive phases varies across regions. For instance, Pennsylvania's reopening plan involves a "red" (initial) phase, a "yellow" phase, and a (terminal) "green" phase. The state's reopening process largely involves each of the 67 counties passing through the three phases. Other states (e.g., New York) have grouped counties into regions that move through the reopening process in tandem.[2] For seven states – Illinois, Michigan, New York, Oregon, Pennsylvania, Virginia, and Washington – we have compiled two worksheets: the movement of counties into different phases, and the list of industries that can be open in each county.

For the second set of states, we instead directly list the NAICS industries that are re-opened on each date. Even for this second group of states, there are certain instances of within-state across-county variation in when industries reopen. For counties that are exceptions to the rest of their respective states, we list these exceptions in our excel file as well. The file that collects information on reopening for these states is `reopening_other.xlsx` . Within this file, there are four worksheets. The first worksheet includes the main information that will be read in by STATA. The second and third worksheets list the counties (and associated FIPS county codes) and the titles of industries (and associated NAICS industry codes). The final worksheet provides a partial list of the sources for this worksheet.

Regarding the mapping between text and NAICS industries, there are three exceptional cases to keep in mind. First, certain states issued blanket reopening orders as part of their reopening. For instance, on June 1 Oklahoma entered "phase 3" of its reopening plan, whereby all industries are allowed to operate (albeit with certain restrictions). In cases like these, we assign all industries to be open. Second, certain states have indicated that "office-based businesses" (or similar phrases) are allowed to reopen. (An example of this would be Rhode Island, in phase 1 of its plan.) In these instances, we take NAICS sectors beginning with "5" (with some exceptions) in addition to NAICS sectors 8132-8139.[3] Third, in their reopening orders, many states mention reopening manufacturing without discussing wholesale and related distributing activities, while other states mention reopening these two sectors in tandem. In our view, it is unlikely that wholesale is still a restricted activity (especially given that manufacturing and retail have been allowed to open in every state.) So, for each state that (i) lists manufacturing as a reopening and (ii) has not, to the current date, explicitly mentioned wholesale as a sector that has re-opened, we designate the re-opening date for wholesale sector (plus distribution support activities and warehousing and storage, NAICS codes 488 and 493 respectively) to be that of the manufacturing sector's reopening date.[4]

---

[2] Within New York, for example, eight counties – Albany, Columbia, Green, Renssalaer, Saratoga, Schenectady, Warren, and Washington – comprise the "Capital District" region. These eight counties moved into phase 1 of New York's plan on May 20, phase 2 on June 3, phase 3 on June 17, and phase 4 on July 1.

[3] The exceptions, within the NAICS 5 sector, are those industries that have been explicitly listed in states' reopening orders. These include 51211 (Motion Picture and Video Production), 51213 (Motion Picture and Video Exhibition), 51912 (Libraries and Archives), 54192 (Photography Services), 54194 (Veterinary Services), and 5617 (Services to Buildings and Dwellings).

[4] Of course, this change only applies to states that had at least partially classified wholesale and relate distribution activities as nonessential in their initial closing orders.

III.B Storing Information in a STATA database

The code to produce a STATA database of states' reopening paths is given in steps 2 to 6 of `create_dataset.do`.

- Step 2 reshapes the excel files corresponding to reopening in Illinois, Michigan, New York, Oregon, Pennsylvania, Virginia, Washington so that it has the same format as all other states.
- Step 3 takes the reshaped files from step 2 and the excel file from other states and stores local macro variables with information on opening dates by counties and industries.
- Step 4 uses the local macro variables from step 3 to construct a database with county-industry pairs as the observation and reopening status (on a weekly basis, from early May to the current period) as different variables. This step also constructs a variable identifying when the industry was closed in its state.
- Step 5 modifies the dataset to account for recent (re-)closures (e.g., Texas' late June closure of bars). The output of this step is `reopening_dataset.dta`
- Step 6 collapses up to a coarser industry definition, the industry definition that allows us to merge to datasets like O*NET and the CPS. The output of this step is `reopening_dataset_combined.dta`

Step 4 takes several hours to run. However, to circumvent running the time-consuming portion of the do file, one may download `reopening_dataset.dta` (from `https://reopeningdata.github.io/reopening_dataset.zip`) and then set the local macro variable skip_step_4_5 equal to 1.

Finally, recently (beginning in the last week of June) certain states have re-imposed restrictions on economic activity. Most of these new closures are targeted to individual industries: bars, movie theaters. However, in other instances (primarily California) restrictions cover a wide swath of industries. For instance, on July 13 California Governor Gavin Newsom ordered nonessential offices in certain "watch list" counties to close. The word "nonessential" refers back to the state's initial designation of essential and nonessential industries. In states' re-closure orders we assume that no industry that was initially classified as essential will be forced to close.

IV. Figures and Tables

We produce the figures and tables in our note and on `https://reopeningdata.github.io/figures.html` using the `reopening_figures.do` do file. In addition to the data on reopening, whose construction we describe above, this do file employs data from the American Time Use Survey (to compute the share of workers who report the ability to work from home in each industry), the County Business Patterns (to compute employment in each industry and county), the Current Population Survey (to compute the age of workers in each industry), O*NET (to compute contact intensity in each occupation), the National Employment Matrix (to compute the number of workers in each industry), and the Quarterly Census of Employment and Wages (to compute the number of workers in each industry). The input files are all collected in in the replication ZIP file.

V. Changes Between Versions

V.A Changes between the first and second versions of our dataset and code

There are four main differences between the first (July 8) and second (July 21) version of our dataset and code. The first three are fixes to errors in our coding of states' reopening orders. The final is an update to include the most recent data.

First, on May 13, 2020 a Wisconsin State Supreme Court decision effectively ended its statewide restrictions on economic activity. In response, individual counties enacted replacement restrictions. We have coded up the restrictions for Dane County, Milwaukee County, and Sauk County. Unfortunately, the code in our July 8 version did not capture the statewide change (Wisconsin minus the three aforementioned counties).

Second, in the first version of our code and database, we had initially assumed that Louisiana had fully reopened by May 15. In reality, the following industries should have marked closed until the week of May 29: Beauty Salons; Nail Salons; and Other Personal Care Services. In addition, the following industries are still closed: Child Daycare Services; Promoters of Performing Arts, Sports; and Amusement, Gambling and Recreation. We have updated our code and dataset to reflect this reality.

Third, within create_dataset.do, there are the following two lines of code:

```
sort statefips countyfips  naics month day
by statefips countyfips  naics: keep if _n==1
```

The idea, here, is to make sure that each NAICS industry only appears once for a given state-county (and that we take the first mention of the NAICS industry as the date of reopening). Within the July 8 version of our do file, the following line of code was missing.

```
sort statefips countyfips  month day naics
```

To see why this line of code is necessary, consider a hypothetical state-county which opens restaurants on May 1 and bars on May 15. In the excel spreadsheet, we would have NAICS=7225 open on May 1 and NAICS=722 on May 15. Without the final extra line of code described above, the macro variable corresponding to NAICs=722 is read *before* the macro variable for 7225. This is wrong. What will happen is that all of 722 (including 7225) will be reported as closed on May 1 and May 8. When adding the extra line of sorting, 7225 is correctly read as open on May 1 and May 8. The error that we have identified leads us to classify the pace of reopening as slower than it actually was. This error was pertinent for slightly less than 0.1 percent of the state-county-industry triples in our dataset.

Finally, we have added two new variables, indicating whether industries were open as of July 17 or as of July 24. Within New York, New York City moved into the final phase of the state's reopening plan. In the other direction, California has increased the scope of its re-closures. Pennsylvania and New York have closed bars statewide. Nevada and Colorado had restrictions on bars that began before July 10, and that we have now included in our dataset.

V.B. Changes between the second and third versions of our dataset and code

There are three main corrections and two speed enhancements between the second (July 21) and third (August 5) versions of the code. Together, the corrections resulted in changes to 0.34% of observations, with respect to the dates at which industries were reopened.

V.B.1 Corrections:

1. Reformatting month and day variables as numeric within Step 3 of `create_dataset.do`.

After updating the sorting procedure in the second version with

```
sort statefips countyfips  naics month day
```

The following line of code was missing immediately preceding the sort:

```
destring month day, replace force
```

Without this line of code, month and day would string variables, meaning that, for example, May 22 as erroneously sorted before May 9.

2. Add an if-condition (month==.) to reopening date replacements.

Reopening dates are associated with either one, two, three, four, five, or six digit NAICS industry codes. We assume that coarser codes map onto several finer codes, if no finer reopening date exists. So, the reopening date associated with 448, clothing and accessory stores in Worcester County, Maryland should become the reopening date for finer NAICS 448320 (leather and luggage stores), 448150 (tie shops), 448210 (shoe stores), since they are within the category of clothing and accessory stores and were not mentioned explicitly within Worcester County's reopening order.

However, in Worcester County industry code 448320 was coded as reopening on June 19, 2020. This date originates from the reopening date for the 2-digit NAICS industry 44, or retail overall. But, in this county, there exists a finer 3-digit NAICS 448 for clothes and accessories with a reopening date of May 15, 2020, a month before all of retail. The sorting in section V.A. orders these two NAICS codes properly as follows:

	NAICS	MONTH	DAY
1.	448	5	15
2.	44	6	19

But the line

```
replace open_may_21 = (date< if substr(naics,1,`qq')==`"naics_`x'"' & fipst==`state_`x'' & `county_`x''==fipsc
```

first sets open_may_21 to 1, from row 1 in the table above. Then row 2 flips open_may_21 to 0 in the subsequent iteration due to the later reopening date.

To resolve this issue, the phrase in bold is added to the end of the line of code

```
replace open_may_21 = (date< if substr(naics,1,`qq')==`"naics_`x'"' & fipst==`state_`x'' & `county_`x''==fipsc & open_may_21==.
```

which in this example allows information to be written from the first row, but skips the second row of the above table. Thus, the reopening date for all of retail will only be written to 6-digit industries starting with 44 that have not yet been filled by finer NAICS codes. In general, this addition prevents overwriting reopening dates from finer NAICS codes with those of coarser codes.

3. Merge reopening dates to (state, county, six-digit NAICS code) triples, in place of NAICS codes of various digits. At the beginning of the step 4_5 loop within `create_dataset.do`, we replace

```
use naics_essential_list, replace
```

with

```
use naics_list, replace
```

The dataset records 1,057 6-digit NAICS codes, resulting in 3,371,830 county-state-NAICS triples when crossed with all 3,190 U.S. counties. Previously, naics_essential_list comprised 604 NAICS codes of various digits that aligned with each industry mentioned in counties' initial essentiality lists, resulting in 1.9M observations when crossed with all U.S. counties. However, there were some cases in which reopening orders were more finely coded, i.e. required more digits, than their initial essentiality designations.

For example, Winchester County, VA initially closed the performing arts sector with and without facilities (NAICS 7113). The then county reopened performing arts without facilities (71132) on June 5, 2020, and the remainder of performing arts (711) on July 1, 2020. Therefore, the line of code:

```
substr("`naics_`x'"",1, qq2)==naics | substr(naics,1,`qq')=="`naics_`x'""
```

was needed to match longer reopening NAICS codes to their shorter counterparts when not present in the 604 industries of naics_essential_list. In this case, this technique matches the reopening date from performing arts without facilities to all performing arts, which is incorrect.

To resolve this issue, we expand to the 6-digit naics_list and remove the first half of the or-statement in the previous line of code, replacing it with:

```
substr(naics,1,`qq')=="`naics_`x'""
```

<u>V.B.2 Speed enhancements:</u>

1. Partition the local macros into 2 lists.

In the most recent version of the code, within the end of Step 3 of `create_dataset.do`, we partition the original local macro lists based on whether the variable `countyfips' is missing. countyfips determines whether a certain industry's reopening date is recorded at the state or the county-state level. Because observations in our reopening_dataset are ultimately county-state-NAICS triples, this partitioning is necessary to define whether reopening dates are county-specific and should map to one observation, or should be written to several observations, counties without county-specific orders within a state.

a. After partitioning, we define two sets of local macro variables: `county_1', `county_2', … and `county2_1', `county2_2',… for county-state and state level reopening orders respectively.

b. Rather than iterating over the all pooled county and state-level orders twice, partitioning the orders beforehand results substantially fewer iterations.

2. Generate open_may_1, open_may_7, … , open_aug_7 variables outside of the macro loop.

V.B.3 Updates to reopening orders:

Note, these updates are also listed at `https:/reopeningdata.github.io/updates`

- We have added variables indicating whether industries were open as of July 31 or as of August 7.
- We have added movie theaters (open at 50% capacity) and gyms to Washington's Phase III essential industry list
- Kentucky closed bars in the last week of July. Santa Cruz County and San Mateo County in California were added to the state's watch list in that week.
- South Carolina has movie theaters, concert halls, and sport arenas to open, as of August 3.
- South Carolina had allowed retail to re-open in late April, something that we had previously missed.
- Previously, we had categorized Allegheny County, Pennsylvania as having closed restaurants and bars from the weeks of July 3 to July 17. In fact, outdoor service was allowed beginning July 8.